Mitchell M. Holland, Ph.D.
Associate Professor, Biochem & MolBio
Former Director, Forensic Science Program
Penn State University, University Park, PA

The Basics of Next Generation DNA Sequencing

MN BCA Workshop
1 June 2015

www.forensics.psu.edu
Topics to Address

- Current State/Impact of NGS
- Benchtop NGS Instruments
- The Pipeline: Input DNA, Library & Template Preparation
- Sequencing Chemistries
- Applications
The Pace of Science

- There are >50,000 human genomes sequenced, to date
At $5000 per human genome, 50,000 genomes would cost $250M

Therefore, we can now sequence 200,000 human genomes for the cost of the first genome sequence

NIH News April 2013
The illumina HiSeq X Ten

Can sequence ~16 human genomes in ~3 days for <$1000 per genome with ~30 X coverage* per nucleotide

The cost covers reagents, labor and instrument depreciation, but not overhead-related expenses

*Coverage is synonymous with RFUs

~$1,000,000 per instrument
Must purchase 10 instruments at one time ($10,000,000). The suite of instruments can produce >18,000 human genome sequences per year

... or >1.5\times10^{15} (1.5 quadrillion) bases of DNA sequence per year, or 10^{-8} cents/nucleotide/yr
NGS Platforms

The “Benchtop” Players

MiSeq
Late 2011

GS Junior
Early 2010

Ion PGM
Early 2011
Instrument Choice

Highest Throughput = 15 Gb/run
Lowest Error Rate

Moderate Throughput = Up to 2 Gb/run
Fastest Output = 275 Mb/hour

Lowest Throughput = 35 Mb/run
Longest Reads = ~400 bp

Gb = gigabases of sequence
The Pipeline

- Make a “Library” of DNA Templates
- Prep (Copy) the Individual DNA Templates for Sequencing
- Sequence the Copied Templates

Input DNA Amplicon, Whole Genome, RNA

CONSTRUCT LIBRARY Includes PCR
PREPARE TEMPLATE Includes PCR
RUN SEQUENCE Sequencing By Synthesis
ANALYZE DATA

Figure from LifeTechnologies
Input DNA

- For example,
  - Whole mtgenome amplification
    - Achieved as either large (~8.5 kb) or small (~400 bp) amplicons
  - Targeted D-loop sequences
    - Achieved as the entire D-loop (~1100 bp) or traditional primer set (~250 bp) amplicons
Library Preparation

enzymatic shearing, ligation of adapters and size selection

Ion Xpress™ Plus Fragment Library Kit

Ion Shear™ enzyme-based DNA fragmentation system removes the need for physical shearing

For amplicons, Ion Plus Library & Barcode Kits to add adaptors and barcodes to individual samples

LifeTechnologies
The Library Builder™ System automates the steps from shearing to barcoding for up to 39 libraries per day.

Highlights the need for automated, high-throughout solutions for the library prep process.

LifeTechnologies
Template Prep: Emulsion PCR & Chip Loading

The Ion Chef™ System:
Amplification mix preparation, emulsification, amplification, breaking, enrichment, washing, sequencing prep, and chip preparation & loading

LifeTechnologies
Sequencing Chemistry

Sequencing By Synthesis
Solid-State pH Meter

LifeTechnologies
Hydrogen ions will change the pH of the solution, which will be detected by the ion sensor, directly translating the chemical signal into a digital format.
The nucleotide does not compliment the template - no release of hydrogen.

The nucleotide compliments the template - hydrogen is released.

The nucleotide compliments several bases in a row - multiple hydrogen ions are released.

Sequential flood of dNTP
Library Preparation
ntagmentation and indexing

Nextera®XT Sample Prep

illuminna
STEP 1: Tagmentation

A transposase cuts the DNA and inserts primer adapters for downstream PCR and sequencing

Nextera®XT Sample Prep

illumina

Library Prep
STEP 2: Indices & Bridge Amplification Adaptors

The inserted primer adapters are used for PCR to incorporate sample identifier indices and bridge amplification adaptors.

Gaps are filled in early in the PCR process.
Template Preparation

bridge amplification

- Construct Library
- Prepare Template
- Run Sequence
- Analyze Data

illumina
Bridge Amp on the Flow Cell

Bridge Amplification Cluster Generation

illuminan
Bridge Amplification (BA)

A single copy of ssDNA attaches to the lawn via P5 or P7 to a complimentary oligomer. A copy of the ssDNA template is made and covalently attaches to the flow cell.
Bridge Amplification
The newly formed copy will bridge over and bind to a complimentary oligomer (P5 or P7), and allow for repeated amplification and cluster formation.
Bridge Amplification

The P5 or P7 strands are selected for by cleaving off the other strands from the lawn, leaving behind a cluster of DNA strands in the same orientation.
All four fluorescecently labeled ddNTPs (terminators) are added together, one is incorporated at each cluster and read by a CCD camera.
Reversible Terminators

Once incorporated and read by the CCD camera, the dye is removed from the terminator along with the protective group on the 3’-hydroxyl group.

illuminax
Paired End Reads
NGS in Forensic Science

Questions to address when introducing NGS in forensic DNA labs

- Applications
- Target Loci
- Technology
- Technology Transfer
- Bioinformatics
- Legal Questions
• Morphological SNP Markers
  • Geoprofiling, eye/hair color, skin pigmentation, etc

• Kinship SNP Markers
High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform

Sarah L. Fordyce, Maria C. Ávila-Arco, Eszter Rockenbauer, Claus Borsting, Rune Frank-Hansen, Frederik Torp-Petersen, Eske Willerslev, Anders J. Hansen, Niels Morling, and M. Thomas P. Gilbert

1 Centre for GeoGenetics, Natural History Museum of Denmark, Copenhagen, Denmark and 2 Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark

Short-read, high-throughput sequencing technology for STR genotyping

Daniel M. Bornman, Mark E. Hester, Jared M. Schetter, Manjula D. Kasoji, Angela Minard-Smith, Curt A. Barden, Scott C. Nelson, Gene D. Godbold, Christine H. Baker, Boyu Yang, Jacqueline E. Walther, Ivan E. Toomes, Pearl S. Yan, Benjamin Rodriguez, Ralf Bundschuh, Michael L. Dickens, Brian A. Young, and Seth A. Faith

1 Battelle Memorial Institute, Columbus, OH, USA, 2 Battelle Memorial Institute, Charlottesville, VA, USA, 3 Human Cancer Genetics Program, The Ohio State University Comprehensive Cancer Center, Columbus, OH, USA, and 4 Department of Physics and Biochemistry, Center for RNA Biology, The Ohio State University, Columbus, OH, USA

EVALUATION OF STANDARD REFERENCE SAMPLE TYPES FOR NEXT GENERATION SEQUENCE-BASED GENOTYPING

Seth A. Faith, Mark E. Hester, Dan M. Bornman, Angela Minard-Smith, Brian A. Young

Battelle Memorial Institute, 505 King Avenue, Columbus, Ohio 43201, USA
Mixture Deconvolution

STRs

Same "Issues" as Normal CE Analysis
Is this the best target for introducing NGS into crime labs & the legal system??